

RESEARCH ARTICLE

Open Access

Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug

Hee Sook Lee^{1,6†}, Taejeong Bae^{1,2,6†}, Ji-Hyun Lee^{1,2,6†}, Dae Gyu Kim¹, Young Sun Oh¹, Yeongjun Jang³, Ji-Tea Kim³, Jong-Jun Lee¹, Alessio Innocenti⁴, Claudiu T Supuran⁴, Luonan Chen⁵, Kyoohyoung Rho^{3*} and Sunghoon Kim^{1,2*}

Abstract

Background: The process of drug discovery and development is time-consuming and costly, and the probability of success is low. Therefore, there is rising interest in repositioning existing drugs for new medical indications. When successful, this process reduces the risk of failure and costs associated with *de novo* drug development. However, in many cases, new indications of existing drugs have been found serendipitously. Thus there is a clear need for establishment of rational methods for drug repositioning.

Results: In this study, we have established a database we call "PharmDB" which integrates data associated with disease indications, drug development, and associated proteins, and known interactions extracted from various established databases. To explore linkages of known drugs to diseases of interest from within PharmDB, we designed the Shared Neighborhood Scoring (SNS) algorithm. And to facilitate exploration of tripartite (Drug-Protein-Disease) network, we developed a graphical data visualization software program called phExplorer, which allows us to browse PharmDB data in an interactive and dynamic manner. We validated this knowledge-based tool kit, by identifying a potential application of a hypertension drug, benzthiazide (TBZT), to induce lung cancer cell death.

Conclusions: By combining PharmDB, an integrated tripartite database, with Shared Neighborhood Scoring (SNS) algorithm, we developed a knowledge platform to rationally identify new indications for known FDA approved drugs, which can be customized to specific projects using manual curation. The data in PharmDB is open access and can be easily explored with phExplorer and accessed via BioMart web service (<http://www.i-pharm.org/>, <http://biomart.i-pharm.org/>).

Keywords: Tripartite network, Drug repositioning, Shared Neighborhood Scoring (SNS) algorithm

Background

Modern drug discovery is time-consuming and expensive, involving coordinated multi-disciplinary research in multiple stages, each requiring intensive and specialized resources [1]. Although rapid advancement of "omics" approaches, computational systems biology and accumulation of digital data resources have provided a vast array of significant information in life science [2], data relevant

to drug discovery are not easily identified and recruited for application to pharmaceutical research [3]. Despite the technological advances in drug discovery such as HTS, the approval of new drugs has remained stagnant in the past decade, resulting in an overall decline in the productivity of the pharmaceutical industry.

In efforts to save development time and minimize the risk of failure during drug development, repositioning of currently available drugs to new therapeutic indications is considered an alternative route [1]. To date most repositioned drugs have been the consequence of serendipitous observations of unexpected efficacy and side effects of drugs in development or on the market. However, recently, systems biology approaches have been applied in efforts to discover unknown effects for existing drugs.

* Correspondence: krho@kribb.re.kr; sungkim@snu.ac.kr

†Equal contributors

³Information Center for Bio-pharmacological Network, Seoul National University, Suwon, Korea

¹Medicinal Bioconvergence Research Center, College of Pharmacy, Seoul National University, Seoul, Korea

Full list of author information is available at the end of the article

For instance, drug repositioning approaches have incorporated *in silico* approaches for analyzing large data sets such as gene expression profiles [4,5], literature mining [6], chemical similarity [7], side-effect similarity [8], disease-drug network [9], pathway-based disease network [10], and phenotypic disease network [11]. To establish a more logical approach to repositioning a known drug to a new indication, we established a knowledge platform comprising binary linkages between diseases, drugs, and proteins, from which new and previously unknown connections can be drawn between drugs and diseases of interest. This integrated database was designated PharmDB.

For probing the database and identifying disease-drug linkages, we have developed the Shared Neighborhood Scoring (SNS) algorithm, which predicts relationships between drugs, proteins and diseases. While the relationship data are collected from experiments, coverage of the data is still incomplete. Thus there may be undetected links and hidden nodes in the network. Up to now, a number of prediction methods and measures have been proposed to find these undetected associations from topological or structural properties of various complex networks [12,13]. To date, most of these algorithms and measures are applicable only to a monopartite network that consists only of one type of node. Therefore, multipartite network composed of more than a type of nodes cannot be analyzed using these measures. To solve this problem, researchers have used projection methods that convert multipartite networks into monopartite ones. Unfortunately, any projection method can result in information loss, especially in low-degree nodes. Accordingly projecting the PharmDB tripartite network into monopartite drug, protein and disease networks can distort many well-known network measures, such as average path length $\langle l \rangle$, average clustering coefficient $\langle C \rangle$, degree-dependent clustering coefficient $C(k)$, degree distribution $P(k)$, assortativity coefficient r [14], and degree-degree correlation coefficient $k_{nn}(k)$ [15]. To overcome

these limits of the projection technique, we designed a new prediction method called Shared Neighborhood Scoring (SNS) algorithm which calculates the probability of a link existence between two nodes of interest. This can be done by evaluating the connections of their neighbors in PharmDB tripartite network.

Results

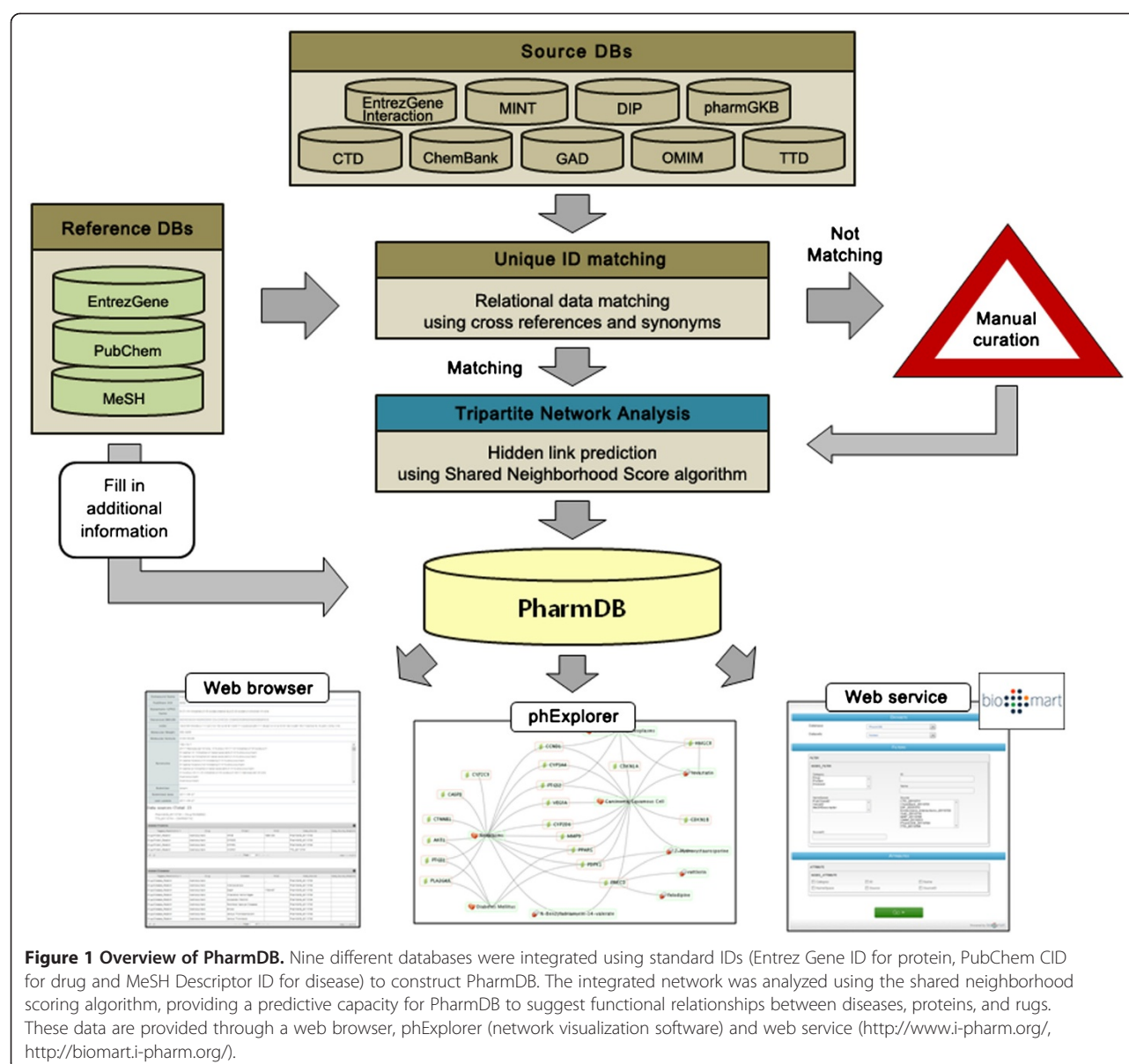
System overview

The PharmDB is a tripartite pharmacological network database consisting of three kinds of nodes: human diseases, FDA approved drugs or druggable chemicals, and proteins. The proteins in PharmDB include therapeutic targets, disease-associated proteins, and drug-metabolizing proteins. The nodes and links used to construct this network database were imported from nine public databases, namely, EntrezGene interaction [16], MINT [17], DIP [18], CTD [19], TTD [20], ChemBank [21], PharmGKB [22], OMIM [23], and GAD [24] (Table 1).

Although these individual databases provide information about the relationships between drugs, diseases, and proteins, they do not provide an integrated network map among the three components in an interactive manner. For data integration in a unified format, we adopted PubChem CID for drugs, GeneID for proteins (tagging separate IDs for isozymes and subunits), and MeSH descriptor for diseases (Figure 1). PharmDB currently includes the nodes of 11,792 drugs, 38,056 proteins, and 6,607 diseases. It also contains 189,800 Drug-Protein, 109,124 Protein-Disease, and 12,232 Drug-Disease, 156,902 Protein-Protein links. The contents of the tripartite pharmacological network in PharmDB are provided through a website (<http://www.i-pharm.org/>). phExplorer, a graphical data visualization software program is also provided for interactive browsing of relevant data. For constructing workflows, PharmDB is provided in BioMart format (<http://biomart.i-pharm.org/>). Currently, software for finding the shortest path between two nodes is only provided through the website.

Table 1 Data sources of PharmDB

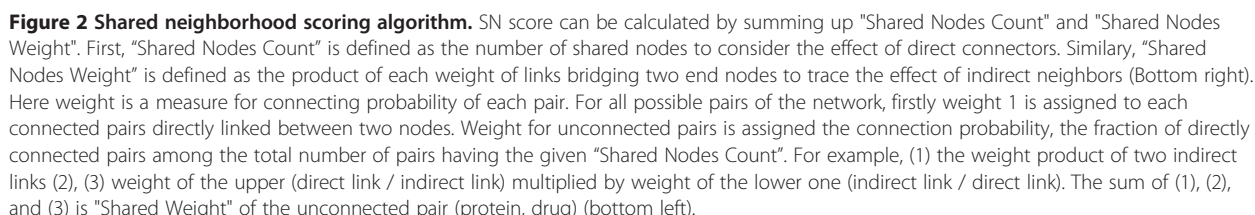
	Drug-Protein Relation	Protein-Disease Relation	Drug-Disease Relation	Protein-Protein Interaction	Update Date
EntrezGene Interaction				V	2011.07.20
MINT				V	2011.07.08
DIP				V	2010.10.10
PharmGKB	V	V	V	V	2011.07.20
CTD	V	V	V		2011.07.11
TTD	V	V	V		2011.07.04
ChemBank			V		2011.07.21
OMIM		V			2011.03.10
GAD		V			2011.07.16



Shared neighborhood scoring (SNS) algorithm

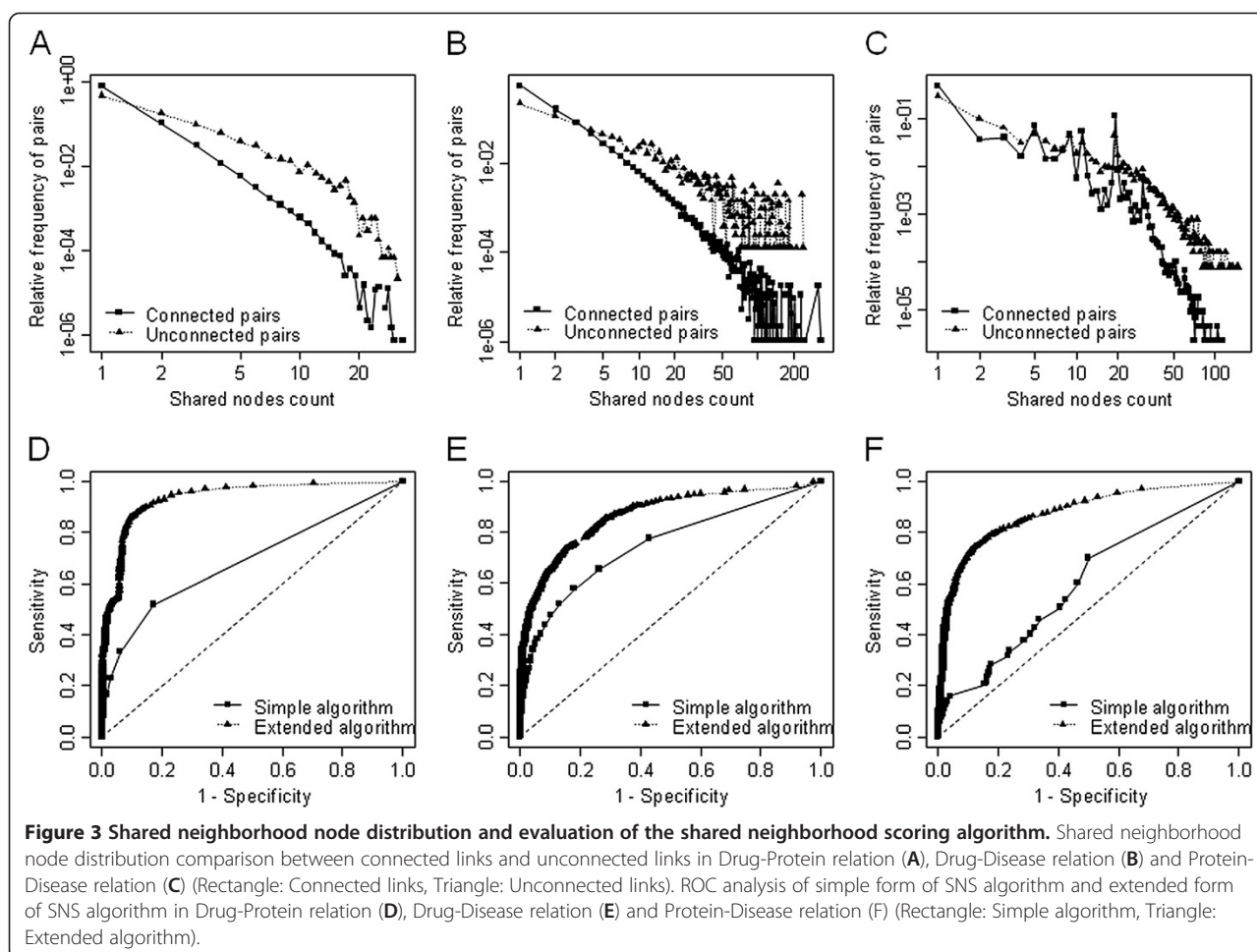
The concept of SNS algorithm is similar to Swanson's ABC model, which applies the transitivity rule to discover missing knowledge from biomedical literature [25]. The SNS algorithm is based on the observation that the probability of connection between two nodes shows monotonic increase with "Shared Nodes Count", the number of in-between nodes connecting two nodes (Figure 2, middle left box). Further we calculated weights for all possible pairs of the network. First we assigned weight 1 to each connected pairs directly linked between two nodes. If a pair of two nodes is not connected, the connection probability is assigned as weight for this indirect link or a virtual link between two nodes. As shown

in the Figure 2, the connection probability for given "Shared Nodes Count" can be computed to be the fraction of directly connected pairs among the total number of pairs having the given "Shared Nodes Count". Finally we developed the share neighborhood score (SN score) by summing up "Shared Nodes Count", the number of shared nodes and "Shared Nodes Weight", the product of each weight of (direct or indirect) links bridging the two end nodes (Figure 2, bottom left). As the SN score possesses a range of values in each relation category (drug-protein, protein-disease, and drug-disease), we developed a normalization method using the connecting probability function of SN score distribution (see Materials and Methods for details).



The prediction performance of the SNS algorithm was measured by plotting receiver operating characteristic (ROC) curves (Figure 3D, 3E, 3F). For calculating SN scores, “simple algorithm” considers only “Shared Nodes Count” but “extended algorithm” includes both “Shared Nodes Count” and “Shared Nodes Weight”. As shown in the Figure 3, the extended algorithm shows better performance than simple one. AUC values of simple algorithm are 0.679, 0.778, and 0.602, in Drug-Protein relation, Drug-Disease relation, and Protein-Disease relation, respectively. And AUC values for extended algorithm are 0.937, 0.868, and 0.871. According to the result, prediction performances with extended scope of shared neighborhood nodes were improved by 38%, 12%, and 45%, respectively.

As a case study, we chose squamous cell carcinoma (SCC) (MeSH descriptor: D002294), a subtype of lung cancer, and tested whether PharmDB could identify any drugs that have a potential for treating this type of cancer. For the primary selection of drug candidates in this case, we made the following criteria. First, they should be inferred by SNS algorithm with SN score bigger than 0.004 and Share Nodes Count zero. Second, they should belong to FDA approved drugs. Third, they should not have been previously used for cancer drug. Forth, they should be directly linked to cancer target proteins (Figure 4). Twenty eight common drugs fit to the four criteria above and were suggested as potential SCC drug candidates (Additional file 1: Table S3). We then went over these candidates to choose the one for experimental validation. Considering technical feasibility, availability of materials, intellectual property and potential for new drug development, we decided to examine thia-benzthiazide (TBZT) whether it can be used for SCC treatment. TBZT is a kind of thiazide

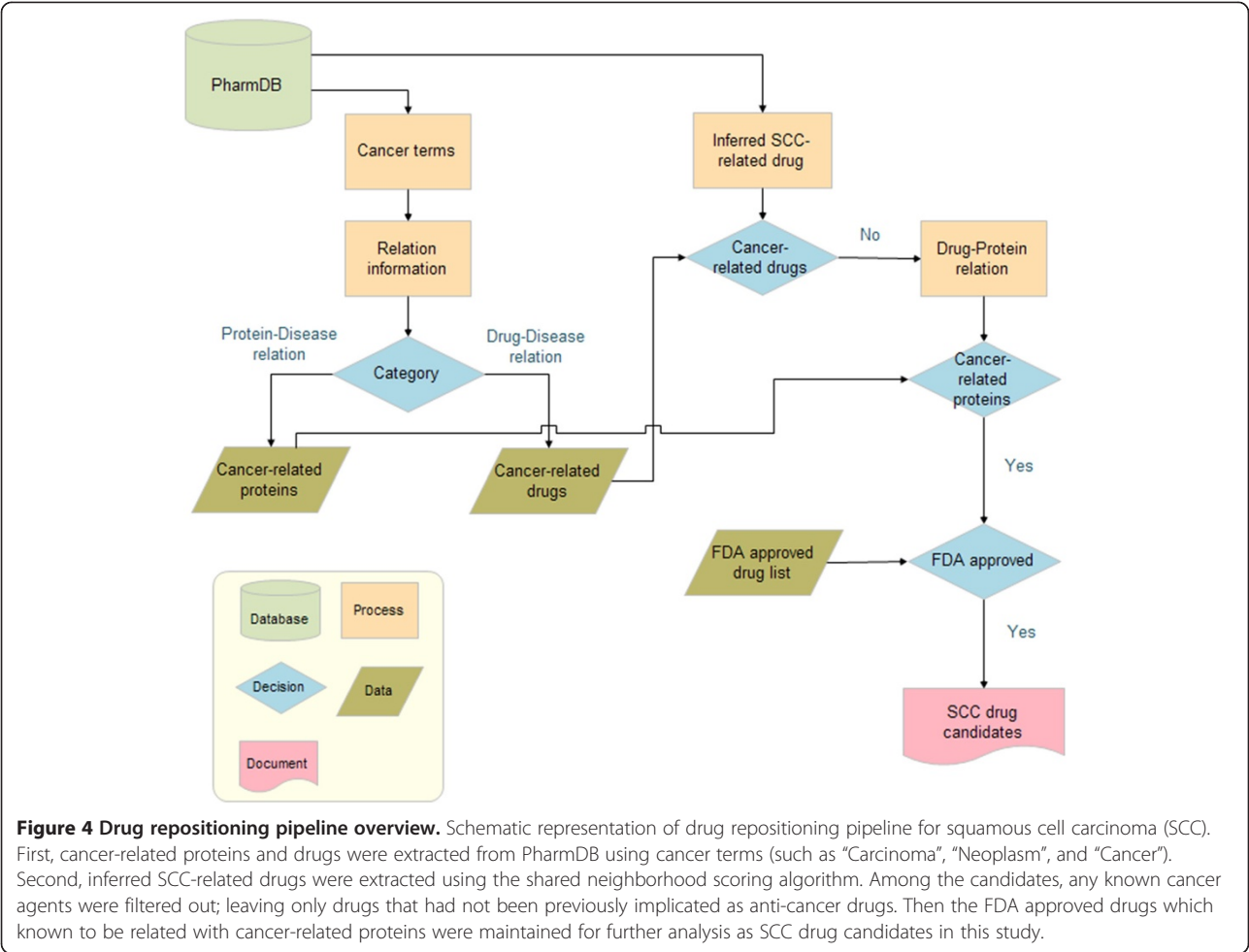


diuretic used for the treatment of high blood pressure and edema [26]. To validate a potential of TBZT as a lung cancer drug, we administered different concentrations of TBZT to squamous lung cancer cells (HCC-1588) under hypoxic conditions (which mimic the tumor microenvironment), as well as under normoxic conditions [27]. Their effects on cell proliferation were monitored by [3 H] thymidine incorporation. We found that under hypoxic conditions only, TBZT can suppress proliferation of cancer cell in a dose-dependent manner (Figure 5A). The hypoxia-dependent cell death induced by TBZT was further confirmed by flow cytometry (Figure 5B).

Carbonic anhydrases (CAs) are zinc metalloenzymes which catalyze the conversion of carbon dioxide to the bicarbonate ion and protons. The CAs are involved in many biological and physical processes including pH homeostasis and have 16 mammalian isoforms (CA1 ~ CA 16) [28]. In PharmDB, TBZT is linked to carbonic anhydrase 2 (CA2). However, TBZT can suppress proliferation of lung cancer cell under hypoxic conditions only (Figure 5) and the expression of CA2 is not

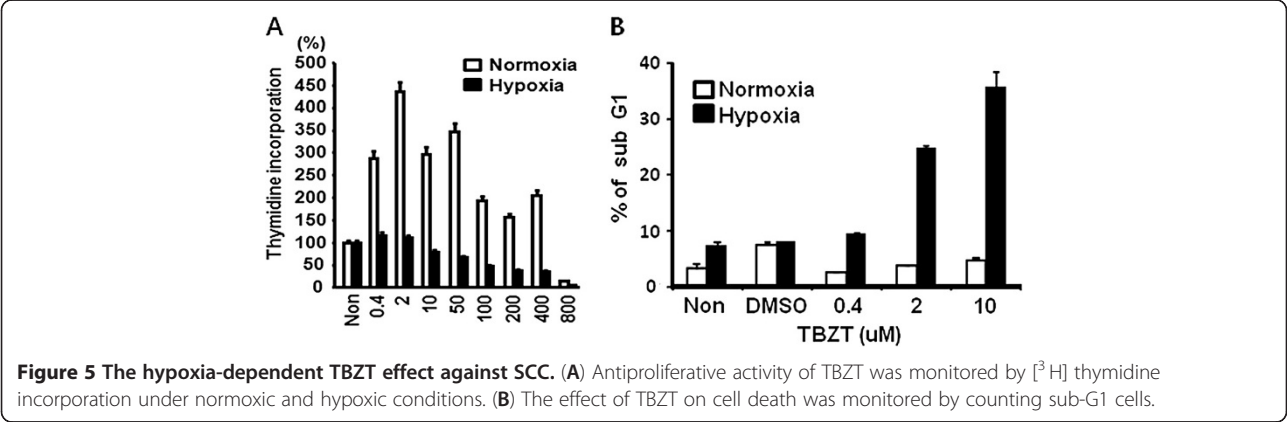
associated with hypoxic conditions. So we have extended cancer-linked CA isoforms in PharmDB (Figure 6A). As a result, we considered three different human CA isozymes (i.e., 1, 2, and 9) as targets of TBZT, and tested whether TBZT inhibits CA activity. TBZT suppressed all of the three CA isozymes with similar K_i values (Figure 6B). As a positive control, acetazolamide (AZA), a known inhibitor of carbonic anhydrases (CAs), was also used [29]. AZA also suppressed the activities of the three CAs, although the K_i values varied depending on the target enzymes. However, among the CA isozymes, CA9 is known to be induced in hypoxic conditions and has functional association with cancer [30]. Thus, the efficacy of TBZT against HCC-1588 cells is likely to have resulted from its inhibition of CA9. For that reasons, we decided to focus on CA9 as the major effective target of TBZT against cancer although we do not exclude the involvement of other isozymes.

CA9, a carbonic anhydrase isoenzyme, is a transmembrane protein that plays an important role in pH regulation [31]. The expression of CA9 is highly induced in various



cancers under hypoxic conditions, which is functionally important for the growth and survival of tumor cells [31]. We confirmed whether CA9 is actually induced in hypoxic conditions by Western blotting with its specific antibody in HCC-1588. As expected, CA9 levels were significantly increased in hypoxic conditions (1% O₂) compared with those in normoxic conditions (20% O₂) (Figure 6C). We also confirmed that TBZT induced cell death by measuring

the activation of caspase 3 (Figure 6D). To confirm the drug-protein pair relationship between CA9 and TBZT, we tested whether the forced expression of CA9 would compensate for the anti-proliferative activity of CA9 by the treatment of TBZT under hypoxic conditions. Cell proliferation was reduced to 70% of the control cells by the treatment of TBZT in the cells transfected with EV, but 35% of the control cells in the cells transfected with CA9.



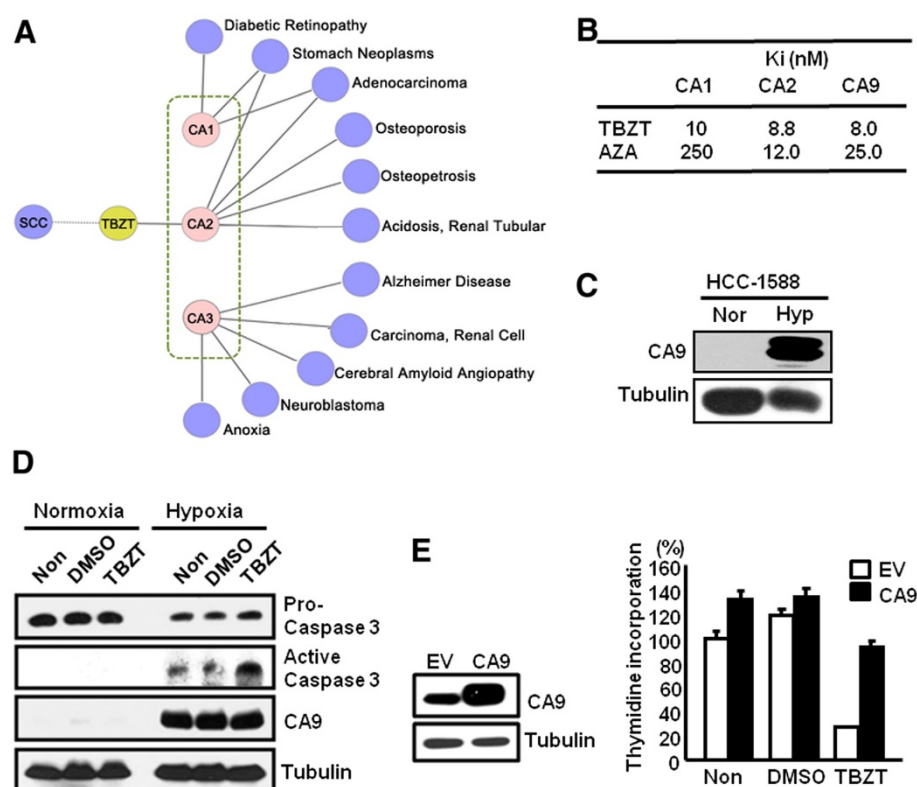


Figure 6 TBZT as an inhibitor of CA9. (A) To validate predictions by PharmDB analysis for SCC, TBZT is tested for inhibitory activity against its potential targets, CA isozymes (CA1, CA2, and CA9). (B) *In vitro* inhibition of TBZT and the AZA control against CA isoforms (i.e., 1, 2, and 9). (C) Cellular levels of CA9 in the SCC cell line, HCC-1588, under normoxic and hypoxic conditions. (D) The effect of TBZT on cell death was monitored by caspase-3 activation. (E) HCC-1588 cells, transfected with an empty vector (EV) or CA9, were treated with TBZT under normoxic and hypoxic conditions.

Therefore, the exogenous supplementation of CA9 recovered the proliferation by up to 35% (Figure 6E). This result validates that the anti-proliferative activity of TBZT against HCC-1588 cells mainly involves CA9. Perhaps, the remaining part could be contributed by other CA isozymes that are also involved in the regulation of cancer. Even if further chemical optimization of TBZT is required to improve efficacy and specificity, these results suggest a possible application of TBZT for further development against lung cancer through its CA9 inhibitory activity.

Discussion

This study demonstrates that drug repositioning can be rapidly guided by a knowledge platform PharmDB, a pharmacological network database comprising protein, drug, and disease data which we are providing as a web-based service. As an ever-increasing amount of biological and pharmacological data are scattered throughout the literature and in proprietary databases, the integrated data of PharmDB provides a valuable tool by consolidating certain valuable sets of data. We adopted a tripartite pharmacological network-based

analysis, and developed a novel neighborhood scoring algorithm to predict previously unknown relationships between drugs, proteins and diseases. The theoretical foundation of algorithm is that a connection probability between two nodes is proportional to the number of nodes commonly shared between them. So the connection probability of two indirectly linked nodes was computed, which is called the shared neighborhood score. This score can highlight missing linkages which may either result from “no actual connection” or “lack of information” and help to differentiate between these two possibilities.

We experimentally validated the usefulness of the shared neighborhood score by identifying a hitherto unknown drug-protein relationship and potential new indication based on this connection. Aside from drug repositioning, the network map of PharmDB composed of not only the data integrated from diverse databases but also the predicted data using the shared neighborhood algorithm can applied to other purposes, such as the prediction of drug mode-of-action, off-target effects, and even the design of optimal drug combinations for a disease of interest.

Conclusions

PharmDB, an integrated tripartite database, coupled with Shared Neighborhood Scoring (SNS) algorithm, would provide much more enriched information than general integrated databases and give us clues for finding new indication of known drugs. Furthermore, these data can be easily explored with phExplorer and accessed via BioMart web service (<http://www.i-pharm.org/>, <http://biomart.i-pharm.org/>).

Methods

Construction of PharmDB

To integrate the data in the existing databases that contain different identifiers, we assigned the following standard identifiers (IDs): PubChem CID for drug, GeneID for protein, and MeSH descriptor for disease. We constructed a comprehensive drug-protein-disease tripartite network by integrating the link information from the nine databases, namely, EntrezGene interaction (<ftp://ftp.ncbi.nih.gov/gene/GeneRIF/>), MINT (<ftp://mint.bio.uniroma2.it/pub/release/mitab26/2011-07-08/>), CTD (<http://ctd.mdibl.org/>), TTD (http://bidd.nus.edu.sg/group/cjttd/TTD_Download.asp), ChemBank (<http://chembank.broadinstitute.org>), DIP (<http://dip.doe-mbi.ucla.edu/dip/Download.cgi>), PharmGKB (http://www.pharmgkb.org/resources/downloads_and_web_services.jsp), OMIM (<ftp://ftp.ncbi.nih.gov/repository/OMIM/ARCHIVE/>), and GAD (<http://geneticassociationdb.nih.gov/>). As the existing databases have their own unique ID systems, we tagged the standard IDs using an in-house script. For the entities that were not tagged by the in-house script, we manually assigned them with appropriate IDs.

Shared neighborhood scoring algorithm

The shared neighborhood scoring algorithm is based on the basic principle that the connection probability of a link between two nodes (i and j) is roughly proportional to the number of nodes commonly shared between the original two nodes, i and j (Additional file 1: Figure S1). The shared neighborhood score S_{ij} is defined as $S_{ij} = \sum_k W_{ik} W_{kj}$. In this equation, i and j indicate the indices of a pair of nodes; k is the index of a shared neighbor node; and W_{ik} is the weight of a link between i and k . The link between i (or j) and k can be real or virtual (i.e., having no known connection but is expected to be connected). Thus, we can define W_{ik} as $W_{ik} = a_{ik} + P(s_{ik}^{(0)})\delta_{a_{ik},0}$. Here, $a_{ik} = 1$ and $\delta_{a_{ik},0} = 0$ if the link between i and k is real; and $a_{ik} = 0$ and $\delta_{a_{ik},0} = 1$ if the link is virtual. When there are only direct connections between node i and node j , a 0th-order shared neighborhood score $s_{ij}^{(0)}$ becomes $s_{ij}^{(0)} = \sum_k a_{ik} a_{kj} = n(0, 1, 2, 3 \dots)$, where n is the number of bridging nodes between node i and node j . $P(s_{ik}^{(0)})$ is a

connection probability that depends on the value of the 0th-order shared neighborhood score $s_{ik}^{(0)}$. For the 0th-order shared neighborhood score $s_{ik}^{(0)} = n(= 0, 1, 2, \dots)$, the function $P(n)$ is defined as follows:

$$P(n) = \frac{(\text{number of connected pairs})}{(\text{number of pairs})} \text{ for } n = 1, 2, \dots$$

$$P(n) = 0 \text{ for } n = 0$$

Based on the probability above, non-linear regression was carried out to extract connecting probability functions. Logistic function was used for this (See Additional file 1: Figure S2, Additional file 1: Table S1 for more details).

$$f(x) = \frac{1}{1 + e^{a+bx}}$$

The shared neighborhood score s_{ij} then becomes $s_{ij} = s_{ij}^{(0)} + s_{ij}^{(1)} + s_{ij}^{(2)} = \sum_k (a_{ik} + P(s_{ik}^{(0)})\delta_{a_{ik},0}) (a_{kj} + P(s_{kj}^{(0)})\delta_{a_{kj},0})$ where $s_{ij}^{(0)} = \sum_k a_{ik} a_{kj}$, $s_{ij}^{(1)} = \sum_k (a_{ik} P(s_{kj}^{(0)})\delta_{a_{kj},0} + a_{kj} P(s_{ik}^{(0)})\delta_{a_{ik},0})$, and $s_{ij}^{(2)} = P(s_{ik}^{(0)})\delta_{a_{ik},0} P(s_{kj}^{(0)})\delta_{a_{kj},0}$. Here, the 1st-order term $s_{ij}^{(1)}$ is added when some nodes are linked directly to node i (or j) but linked indirectly to node j (or i). The 2nd-order term $s_{ij}^{(2)}$ is considered only when some nodes are linked indirectly to both node i and node j . In Additional file 1: Figure S1, as node m is the only shared neighbor of node i and node j , $s_{ij}^{(0)} = 1$. To obtain the 1st-order shared neighborhood score $s_{ij}^{(1)}$ or the 2nd-order shared neighborhood score $s_{ij}^{(2)}$, connection probability $P(s_{ik}^{(0)})$ is calculated beforehand. As a pair (i, k) is mediated by two nodes, $W_{ik} = P(2)$, Path (i, k, j) is composed of an indirect link (i, k) and a direct link (k, j). Similarly, $W_{il} = P(3)$ and $W_{lj} = P(1)$. Path (i, l, j) is composed of both indirect links (i, l) and (l, j). The total shared neighborhood score is thus $s_{ij} = s_{ij}^{(0)} + s_{ij}^{(1)} + s_{ij}^{(2)} = 1 + P(2) + P(3)P(1)$. When calculating $s_{ij}^{(2)}$, we omitted a link between node i and j to remove the dependency of the measure on the existence of a link between i and j , which is the so-called "leave-one-out approach" [32].

The shared neighborhood score is proportional to the number of shared neighborhood nodes. So there is no an upper limit on score. The problem is that the amount of data is not evenly distributed on each relation category. So, even if two different types of relations have identical score, their connecting possibility can't be regarded as identical. For that reason, we normalized the

shared neighborhood score using connecting probability function (Additional file 1: Figure S3).

FDA approved drugs

We have downloaded Drugs@FDA data files (Last updated: 19/09/2011)(<http://www.fda.gov/downloads/Drugs/InformationOnDrugs/UCM163762>). Then we extracted single active ingredient from Product table and tagged PubChem ID for them. The total number of FDA approved drugs tagged with PubChem ID is 23,191.

Cell culture and materials

The HCC-1588 cell line was obtained from the Korean cell line bank and was maintained in RPMI (Hyclone) containing 10% fetal bovine serum and 1% antibiotics. Antibody against caspase-3 and tublin (Cell Signaling Technology) were purchased. M73 monoclonal antibody to CA9 was obtained from Dr. S. Pastorekova (Slovak Academy of Science, Slovak Republic). TBZT and AZA were purchased from Sigma.

Thymidine incorporation assay

To determine the effect of TBZT on cell proliferation, HCC-1588 cells were treated with TBZT in 2% serum-containing media for 48 h under normoxic (20% O₂) and hypoxic (1% O₂) conditions. AZA was used as positive control. pcDNA3-CA9 vector and empty vector (Dr. J.-Y. Kim, National Cancer Center, Korea) were transfected into HCC-1588 cells using Lipofectamine 2000 (Invitrogen). After 24 h incubation, TBZT was added to 2% serum-containing media for 48 h under hypoxic conditions. [³H] thymidine at 1 μCi/ml was added to the culture medium and was incubated for 4 h. The incorporated thymidine was measured by liquid scintillation counter (Wallac).

Flow cytometry

HCC-1588 cells were treated with TBZT (0.4, 2, 10 μM) in 2% serum-containing medium for 48 h under normoxic and hypoxic conditions. AZA was used as positive control. The treated cells were fixed with 70% ethanol for 1 h at 4°C, washed twice with ice-cold PBS, and stained with propidium iodide (50 μg/ml) containing 0.1% sodium citrate, 0.3% NP-40 (nonylphenoxypolyethoxyethanol 40), and 50 μg/ml RNase A for 40 min. The cells were subjected to flow cytometry (FACSCalibur, Becton-Dickinson) to evaluate the apoptotic cells by counting the sub-G1 cells. For each sample, 20,000 cells were analyzed using Cell Quest Pro software.

Enzyme activity

An applied photophysics stopped-flow instrument was used for assaying CA-catalyzed CO₂ hydration activity [33]. Following the initial rates of the CA-catalyzed CO₂

hydration reaction for a period of 10–100 s, phenol red (at a concentration of 0.2 mM) was used as the indicator, working at the absorbance maximum of 557 nm in 20 mM HEPES buffer (pH 7.5) and 20 mM Na₂SO₄ (to maintain the constant ionic strength). For the determination of the kinetic parameters and inhibition constants, the CO₂ concentrations used ranged from 1.7–17 mM. For each inhibitor, at least six traces of the initial 5–10% of the reaction were used for determining the initial velocity. The uncatalyzed rates were determined in the same manner and were subtracted from the total observed rates. Stock solutions of the inhibitor (0.1 mM) were prepared in distilled-deionized water and diluted to 0.01 nM with distilled-deionized water. Inhibitor and enzyme solutions were preincubated together for 15 min–72 h at room temperature (15 min) or 4°C (all other incubation times) prior to assay to allow the formation of the enzyme-inhibitor complex or the eventual active site mediated hydrolysis of the inhibitor. The inhibition constants were obtained by non-linear least-squares methods using PRISM 3 as previously described. The mean values were represented from at least three different determinations [31,34].

Additional file

Additional file 1: Supplementary figures and tables. Detailed Description of Shared neighborhood score algorithm and algorithm Validation. Detailed description of the algorithm including non-linear regression results for extracting connecting probability functions. And a list of inferred SCC drug candidates is included. References.

Abbreviations

HTS: Highthroughput Screening; SNS: Shared Neighborhood Scoring; SN score: Shared Neighborhood score; MINT: the Molecular Interaction database; DIP: the Database of Interacting Proteins; CTD: The Comparative Toxicogenomics Database; TTD: Therapeutic Target Database ChemBank; PharmGKB: The Pharmacogenomics Knowledge Base; OMIM: Online Mendelian Inheritance in Man; GAD: Genetic Association Database; ROC: Receiver Operating Characteristic; AUC: Area Under Curve; AZA: Acetazolamide; CA: Carbonic anhydrase; SCC: Squamous Cell Carcinoma; TBZT: Thia-benzthiazide; KS test: Kolmogorov-Smirnov test.

Competing Interests

The authors declare that they have no competing interests.

Authors' contributions

KR, LC and SK conceived the ideas for this study. KR, JHL, TB and HSL designed and performed the computational systems biology approach. KR, JHL and TB designed and constructed the database. KR and TB developed the algorithm. HSL and JHL designed and performed the repositioning pipeline. HSL, DGK and YSO performed the cell biology experiments. AI and CTS performed the enzyme activity analysis. SK wrote the manuscript with contributions from all the participating authors.

Authors' information

Current positions: KR, YJ - Korean Bioinformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305–806, Korea

Acknowledgments

The authors thank to Sangho Oh and Nam-Jin Koo at Korean Bioinformation Center for updating PharmDB data. This study was supported by the grants of the Global Frontier (NRF-M1AXA002-2010-0029785) and the Research Information Center Supporting Program (2012-0000350) and the WCU project (R31-2008-000-10103-0) of the Ministry of Education, Science, and Technology and Korea Healthcare Technology (A092255-0911-1110100), the Ministry of Health and Welfare Affairs, and Gyeonggi-do to SK, an EU project of the 7th framework programme (METOXIA) to CTS, and by the Korean Ministry of Education, Science and Technology (MEST) under grant number 20110002321.

Author details

¹Medicinal Bioconvergence Research Center, College of Pharmacy, Seoul National University, Seoul, Korea. ²World Class University Program Department of Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul 151-742, Korea. ³Information Center for Biopharmacological Network, Seoul National University, Suwon, Korea. ⁴Dipartimento di Chimica Laboratorio di Chimica Bioinorganica, University of Florence, Via della Lastruccia, 3, Rm. 188 Polo Scientifico, Sesto Fiorentino (Firenze) 50019, Italy. ⁵Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China. ⁶Medicinal Bioconvergence Research Center, Advanced Institutes of Convergence Technology, Suwon 443-270, Korea.

Received: 5 January 2012 Accepted: 31 May 2012

Published: 2 July 2012

References

1. Loding W, Harland L, Williams-Jones B: **High-throughput electronic biology: mining information for drug discovery.** *Nature reviews Drug discovery* 2007, **6**(3):220-230.
2. Butcher EC, Berg EL, Kunkel EJ: **Systems biology in drug discovery.** *Nat Biotechnol* 2004, **22**(10):1253-1259.
3. Wishart DS: **Discovering drug targets through the web.** *Comparative biochemistry and physiology Part D, Genomics & proteomics* 2007, **2**(1):9-17.
4. Lamb J: **The Connectivity Map: a new tool for biomedical research.** *Nat Rev Cancer* 2007, **7**(1):54-60.
5. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, et al: **The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.** *Science* 2006, **313**(5795):1929-1935.
6. Li J, Zhu X, Chen JY: **Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts.** *PLoS computational biology* 2009, **5**(7):e1000450.
7. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kujijer MB, Matos RC, Tran TB, et al: **Predicting new molecular targets for known drugs.** *Nature* 2009, **462**(7270):175-181.
8. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P: **Drug target identification using side-effect similarity.** *Science* 2008, **321**(5886):263-266.
9. Hu G, Agarwal P: **Human disease-drug network based on genomic expression profiles.** *PLoS One* 2009, **4**(8):e6536.
10. Li Y, Agarwal P: **A pathway-based view of human diseases and disease relationships.** *PLoS One* 2009, **4**(2):e4346.
11. Hidalgo CA, Blumm N, Barabasi AL, Christakis NA: **A dynamic network approach for the study of human phenotypes.** *PLoS computational biology* 2009, **5**(4):e1000353.
12. Bailly-Bechet M, Borgs C, Braunstein A, Chayes J, Dagkessamanskaia A, Francois JM, Zecchina R: **Finding undetected protein associations in cell signaling by belief propagation.** *Proc Natl Acad Sci U S A* 2011, **108**(2):882-887.
13. Lu L, Jin CH, Zhou T: **Similarity index based on local paths for link prediction of complex networks.** *Phys Rev E Stat Nonlinear Soft Matter Phys* 2009, **80**(4 Pt 2):046122.
14. Newman ME: **Mixing patterns in networks.** *Phys Rev E Stat Nonlinear Soft Matter Phys* 2003, **67**(2 Pt 2):026126.
15. Pastor-Satorras R, Vazquez A, Vespignani A: **Dynamical and correlation properties of the internet.** *Phys Rev Lett* 2001, **87**(25):258701.
16. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res*, **39**:52-57.
17. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **MINT, the molecular interaction database: 2009 update.** *Nucleic Acids Res* 2010, **38**:532-539. Database issue.
18. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update.** *Nucleic Acids Res* 2004, **32**:449-451. Database issue.
19. Davis AP, King BL, Mockus S, Murphy CG, Saraceni-Richards C, Rosenstein M, Wiegiers T, Mattingly CJ: **The Comparative Toxicogenomics Database: update 2011.** *Nucleic Acids Res* 2011, **39**:1067-1072. Database issue.
20. Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, Huang L, Guo Y, Han L, Zheng C, et al: **Update of TTD: Therapeutic Target Database.** *Nucleic Acids Res* 2010, **38**:787-791. Database issue.
21. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, Brudz S, Sullivan JP, Muhlich J, Serrano M, et al: **ChemBank: a small-molecule screening and cheminformatics resource database.** *Nucleic Acids Res* 2008, **36**:351-359. Database issue.
22. Thorn CF, Klein TE, Altman RB: **Pharmacogenomics and bioinformatics: PharmGKB.** *Pharmacogenomics* 2010, **11**(4):501-505.
23. OMIM (TM): **McKusick-Nathans Institute of Genetic Medicine Online Mendelian Inheritance in Man.** In Bethesda, MD: Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information. Bethesda, MD: National Library of Medicine; 2009. http://www.ncbi.nlm.nih.gov/omim/.
24. Becker KG, Barnes KC, Bright TJ, Wang SA: **The genetic association database.** *Nat Genet* 2004, **36**(5):431-432.
25. Swanson DR, Smalheiser NR: **An interactive system for finding complementary literatures: a stimulus to scientific discovery.** *Artif Intell* 1997, **91**(2):183-203.
26. Havard CW, Wood PH: **Clinical evaluation of benzthiazide, an oral diuretic.** *British medical journal* 1960, **1**(5188):1773-1776.
27. Yotnda P, Wu D, Swanson AM: **Hypoxic tumors and their effect on immune cells and cancer therapy.** *Methods Mol Biol* 2010, **651**:1-29.
28. Robertson N, Potter C, Harris AL: **Role of carbonic anhydrase IX in human tumor cell growth, survival, and invasion.** *Cancer Res* 2004, **64**(17):6160-6165.
29. Xiang Y, Ma B, Li T, Yu HM, Li XJ: **Acetazolamide suppresses tumor metastasis and related protein expression in mice bearing Lewis lung carcinoma.** *Acta pharmacologica Sinica* 2002, **23**(8):745-751.
30. Winum JY, Rami M, Scozzafava A, Montero JL, Supuran C: **Carbonic anhydrase IX: a new druggable target for the design of antitumor agents.** *Medicinal research reviews* 2008, **28**(3):445-463.
31. Supuran CT: **Carbonic anhydrases: novel therapeutic applications for inhibitors and activators.** *Nature reviews Drug discovery* 2008, **7**(2):168-181.
32. Goldberg DS, Roth FP: **Assessing experimentally derived interactions in a small world.** *Proc Natl Acad Sci U S A* 2003, **100**(8):4372-4376.
33. Khalifah RG: **The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C.** *J Biol Chem* 1971, **246**(8):2561-2573.
34. Maresca A, Temperini C, Vu H, Pham NB, Poulsen SA, Scozzafava A, Quinn RJ, Supuran CT: **Non-zinc mediated inhibition of carbonic anhydrases: coumarins are a new class of suicide inhibitors.** *J Am Chem Soc* 2009, **131**(8):3057-3062.

doi:10.1186/1752-0509-6-80

Cite this article as: Lee et al.: Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. *BMC Systems Biology* 2012 **6**:80.